Claims

What is claimed is:

1. A multi-modal conversational computing system, the system comprising:

a user interface subsystem, the user interface subsystem being configured to input multi-modal data from an environment in which the user interface subsystem is deployed, the multi-modal data including data associated with a first modality input sensor and data associated with at least a second modality input sensor, and the environment including one or more users and one or more devices which are controllable by the multi-modal system;

at least one processor, the at least one processor being operatively coupled to the user interface subsystem and being configured to: (i) receive at least a portion of the multi-modal input data from the user interface subsystem; (ii) make a determination of at least one of an intent, a focus and a mood of at least one of the one or more users based on at least a portion of the received multi-modal input data; and (iii) cause execution of one or more actions to occur in the environment based on at least one of the determined intent, the determined focus and the determined mood; and

memory, operatively coupled to the at least one processor, which stores at least a portion of results associated with the intent, focus and mood determinations made by the processor for possible use in a subsequent determination.

2. The system of claim 1, wherein the intent determination comprises resolving referential ambiguity associated with the one or more users in the environment based on at least a portion of the received multi-modal data.

3. The system of claim 1, wherein the intent determination comprises resolving referential ambiguity associated with the one or more devices in the environment based on at least a portion of the received multi-modal data.

4. The system of claim 1, wherein the execution of one or more actions in the environment comprises controlling at least one of the one or more devices in the environment to at least one of effectuate the determined intent, effect the determined focus, and effect the determined mood of the one or more users.

5. The system of claim 1, wherein the execution of one or more actions in the environment comprises controlling at least one of the one or more devices in the environment to request further user input to assist in making at least one of the determinations.

6. The system of claim 1, wherein the execution of the one or more actions comprises initiating a process to at least one of further complete, correct, and disambiguate what the system understands from previous input.

7. The system of claim 1, wherein the at least one processor is further configured to abstract the received multi-modal input data into one or more events prior to making the one or more determinations.

8. The system of claim 1, wherein the at least one processor is further configured to perform one or more recognition operations on the received multi-modal input data prior to making the one or more determinations.

9. A multi-modal conversational computing system, the system comprising:

a user interface subsystem, the user interface subsystem being configured to input multi-modal data from an environment in which the user interface subsystem is deployed, the multi-modal data including data associated with a first modality input sensor and data associated with at least a second modality input sensor, and the environment including

one or more users and one or more devices which are controllable by the multi-modal system;

an input/output manager module operatively coupled to the user interface subsystem and configured to abstract the multi-modal input data into one or more events;

one or more recognition engines operatively coupled to the input/output manager module and configured to perform, when necessary, one or more recognition operations on the abstracted multi-modal input data;

a dialog manager module operatively coupled to the one or more recognition engines and the input/output manager module and configured to: (i) receive at least a portion of the abstracted multi-modal input data and, when necessary, the recognized multi-modal input data; (ii) make a determination of an intent of at least one of the one or more users based on at least a portion of the received multi-modal input data; and (iii) cause execution of one or more actions to occur in the environment based on the determined intent;

a focus and mood classification module operatively coupled to the one or more recognition engines and the input/output manager module and configured to: (i) receive at least a portion of the abstracted multi-modal input data and, when necessary, the recognized multi-modal input data; (ii) make a determination of at least one of a focus and a mood of at least one of the one or more users based on at least a portion of the received multi-modal input data; and (iii) cause execution of one or more actions to occur in the environment based on at least one of the determined focus and mood; and

a context stack memory operatively coupled to the dialog manager module, the one or more recognition engines and the focus and mood classification module, which stores at least a portion of results associated with the intent, focus and mood determinations made by the dialog manager and the classification module for possible use in a subsequent determination.

10. A computer-based conversational computing method, the method comprising the steps of:

obtaining multi-modal data from an environment including one or more users and one or more controllable devices, the multi-modal data including data associated with a first modality input sensor and data associated with at least a second modality input sensor;

making a determination of at least one of an intent, a focus and a mood of at least one of the one or more users based on at least a portion of the obtained multi-modal input data;

causing execution of one or more actions to occur in the environment based on at least one of the determined intent, the determined focus and the determined mood; and

storing at least a portion of results associated with the intent, focus and mood determinations for possible use in a subsequent determination.

11. The method of claim 10, wherein the intent determination step comprises resolving referential ambiguity associated with the one or more users in the environment based on at least a portion of the received multi-modal data.

12. The method of claim 10, wherein the intent determination step comprises resolving referential ambiguity associated with the one or more devices in the environment based on at least a portion of the received multi-modal data.

13. The method of claim 10, wherein the step of causing the execution of one or more actions in the environment comprises controlling at least one of the one or more devices in the environment to at least one of effectuate the determined intent, effect the determined focus, and effect the determined mood of the one or more users.

14. The method of claim 10, wherein the step of causing the execution of one or more actions in the environment comprises controlling at least one of the one or more devices in the environment to request further user input to assist in making at least one of the determinations.

15. The method of claim 10, wherein the step of causing the execution of the one or more actions comprises initiating a process to at least one of further complete, correct, and disambiguate what the system understands from previous input.

16. The method of claim 10, wherein further comprising the step of abstracting the received multi-modal input data into one or more events prior to making the one or more determinations.

17. The method of claim 10, further comprising the step of performing one or more recognition operations on the received multi-modal input data prior to making the one or more determinations.

18. An article of manufacture for performing conversational computing, comprising a machine readable medium containing one or more programs which when executed implement the steps of:

obtaining multi-modal data from an environment including one or more users and one or more controllable devices, the multi-modal data including data associated with a first modality input sensor and data associated with at least a second modality input sensor;

making a determination of at least one of an intent, a focus and a mood of at least one of the one or more users based on at least a portion of the obtained multi-modal input data;

causing execution of one or more actions to occur in the environment based on at least one of the determined intent, the determined focus and the determined mood; and

storing at least a portion of results associated with the intent, focus and mood determinations for possible use in a subsequent determination.

19. A multi-modal conversational computing system, the system comprising:

a user interface subsystem, the user interface subsystem being configured to input multi-modal data from an environment in which the user interface subsystem is deployed, the multi-modal data including at least audio-based data and image-based data, and the environment including one or more users and one or more devices which are controllable by the multi-modal system;

at least one processor, the at least one processor being operatively coupled to the user interface subsystem and being configured to: (i) receive at least a portion of the multi-modal input data from the user interface subsystem; (ii) make a determination of at least one of an intent, a focus and a mood of at least one of the one or more users based on at least a portion of the received multi-modal input data; and (iii) cause execution of one or more actions to occur in the environment based on at least one of the determined intent, the determined focus and the determined mood; and

memory, operatively coupled to the at least one processor, which stores at least a portion of results associated with the intent, focus and mood determinations made by the processor for possible use in a subsequent determination.

20. The system of claim 19, wherein the intent determination comprises resolving referential ambiguity associated with the one or more users in the environment based on at least a portion of the received multi-modal data.

21. The system of claim 19, wherein the intent determination comprises resolving referential ambiguity associated with the one or more devices in the environment based on at least a portion of the received multi-modal data.

22. The system of claim 19, wherein the user interface subsystem comprises one or more image capturing devices, deployed in the environment, for capturing the image-based data.

23. The system of claim 22, wherein the image-based data is at least one of in the visible wavelength spectrum and not in the visible wavelength spectrum.

24. The system of claim 22, wherein the image-based data is at least one of video, infrared, and radio frequency-based image data.

25. The system of claim 19, wherein the user interface subsystem comprises one or more audio capturing devices, deployed in the environment, for capturing the audio-based data.

26. The system of claim 25, wherein the one or more audio capturing devices comprise one or more microphones.

27. The system of claim 19, wherein the user interface subsystem comprises one or more graphical user interface-based input devices, deployed in the environment, for capturing graphical user interface-based data.

28. The system of claim 19, wherein the user interface subsystem comprises a stylus-based input device, deployed in the environment, for capturing handwritten-based data.

29. The system of claim 19, wherein the execution of one or more actions in the environment comprises controlling at least one of the one or more devices in the environment to at least one of effectuate the determined intent, effect the determined focus, and effect the determined mood of the one or more users.

30. The system of claim 19, wherein the execution of one or more actions in the environment comprises controlling at least one of the one or more devices in the environment to request further user input to assist in making at least one of the determinations.

31. The system of claim 19, wherein the at least one processor is further configured to abstract the received multi-modal input data into one or more events prior to making the one or more determinations.

32. The system of claim 19, wherein the at least one processor is further configured to perform one or more recognition operations on the received multi-modal input data prior to making the one or more determinations.

33. The system of claim 32, wherein one of the one or more recognition operations comprises speech recognition.

34. The system of claim 32, wherein one of the one or more recognition operations comprises speaker recognition.

35. The system of claim 32, wherein one of the one or more recognition operations comprises gesture recognition.

36. The system of claim 19, wherein the execution of the one or more actions comprises initiating a process to at least one of further complete, correct, and disambiguate what the system understands from previous input.

37. A multi-modal conversational computing system, the system comprising:

a user interface subsystem, the user interface subsystem being configured to input multi-modal data from an environment in which the user interface subsystem is deployed, the multi-modal data including at least audio-based data and image-based data, and the environment including one or more users and one or more devices which are controllable by the multi-modal system;

an input/output manager module operatively coupled to the user interface subsystem and configured to abstract the multi-modal input data into one or more events;

one or more recognition engines operatively coupled to the input/output manager module and configured to perform, when necessary, one or more recognition operations on the abstracted multi-modal input data;

a dialog manager module operatively coupled to the one or more recognition engines and the input/output manager module and configured to: (i) receive at least a portion of the abstracted multi-modal input data and, when necessary, the recognized multi-modal input data; (ii) make a determination of an intent of at least one of the one or more users based on at least a portion of the received multi-modal input data; and (iii) cause execution of one or more actions to occur in the environment based on the determined intent;

a focus and mood classification module operatively coupled to the one or more recognition engines and the input/output manager module and configured to: (i) receive at least a portion of the abstracted multi-modal input data and, when necessary, the recognized multi-modal input data; (ii) make a determination of at least one of a focus and a mood of at least one of the one or more users based on at least a portion of the

received multi-modal input data; and (iii) cause execution of one or more actions to occur in the environment based on at least one of the determined focus and mood; and

a context stack memory operatively coupled to the dialog manager module, the one or more recognition engines and the focus and mood classification module, which stores at least a portion of results associated with the intent, focus and mood determinations made by the dialog manager and the classification module for possible use in a subsequent determination.

38. A computer-based conversational computing method, the method comprising the steps of:

obtaining multi-modal data from an environment including one or more users and one or more controllable devices, the multi-modal data including at least audio-based data and image-based data;

making a determination of at least one of an intent, a focus and a mood of at least one of the one or more users based on at least a portion of the obtained multi-modal input data;

causing execution of one or more actions to occur in the environment based on at least one of the determined intent, the determined focus and the determined mood; and

storing at least a portion of results associated with the intent, focus and mood determinations for possible use in a subsequent determination.